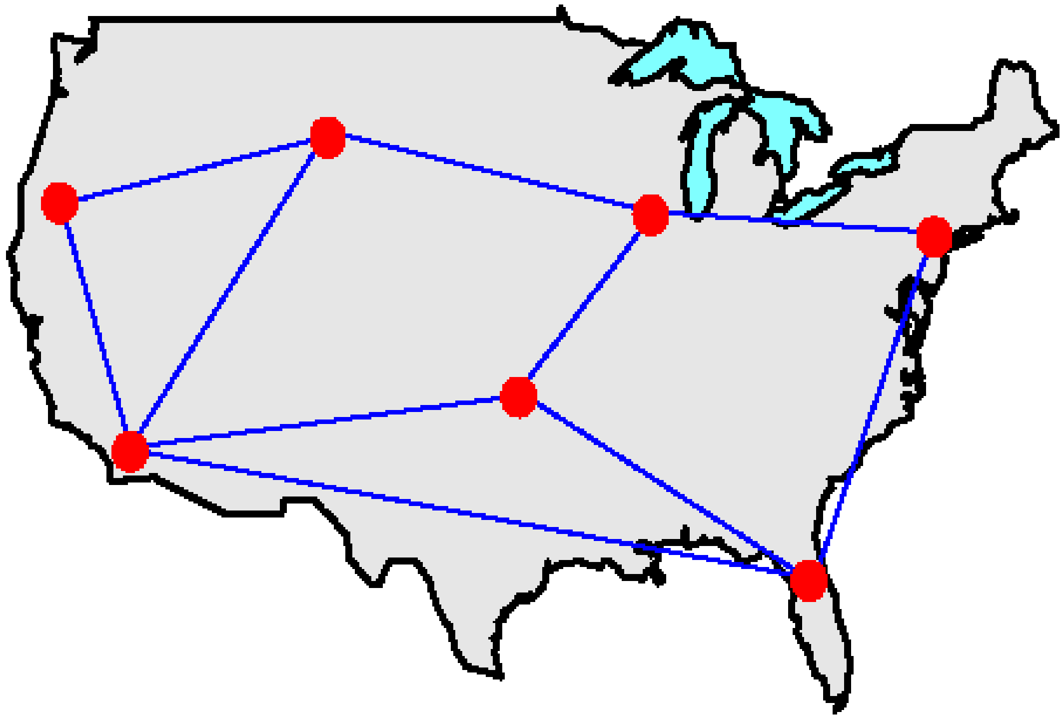
Global Data Grids for 21st Century Science

Paul Avery
University of Florida
http://www.phys.ufl.edu/~avery/
avery@phys.ufl.edu

Global Accelerator Network Workshop
Shelter Island, LI
Sept. 17–20, 2002

The Grid Concept

- ➔Grid: Geographically distributed computing resources configured for coordinated use
  - ◆Fabric: Physical resources & networks provide raw capability
  - ◆Middleware: Software ties it all together (tools, services, etc.)
- ➔Goal: Transparent resource sharing

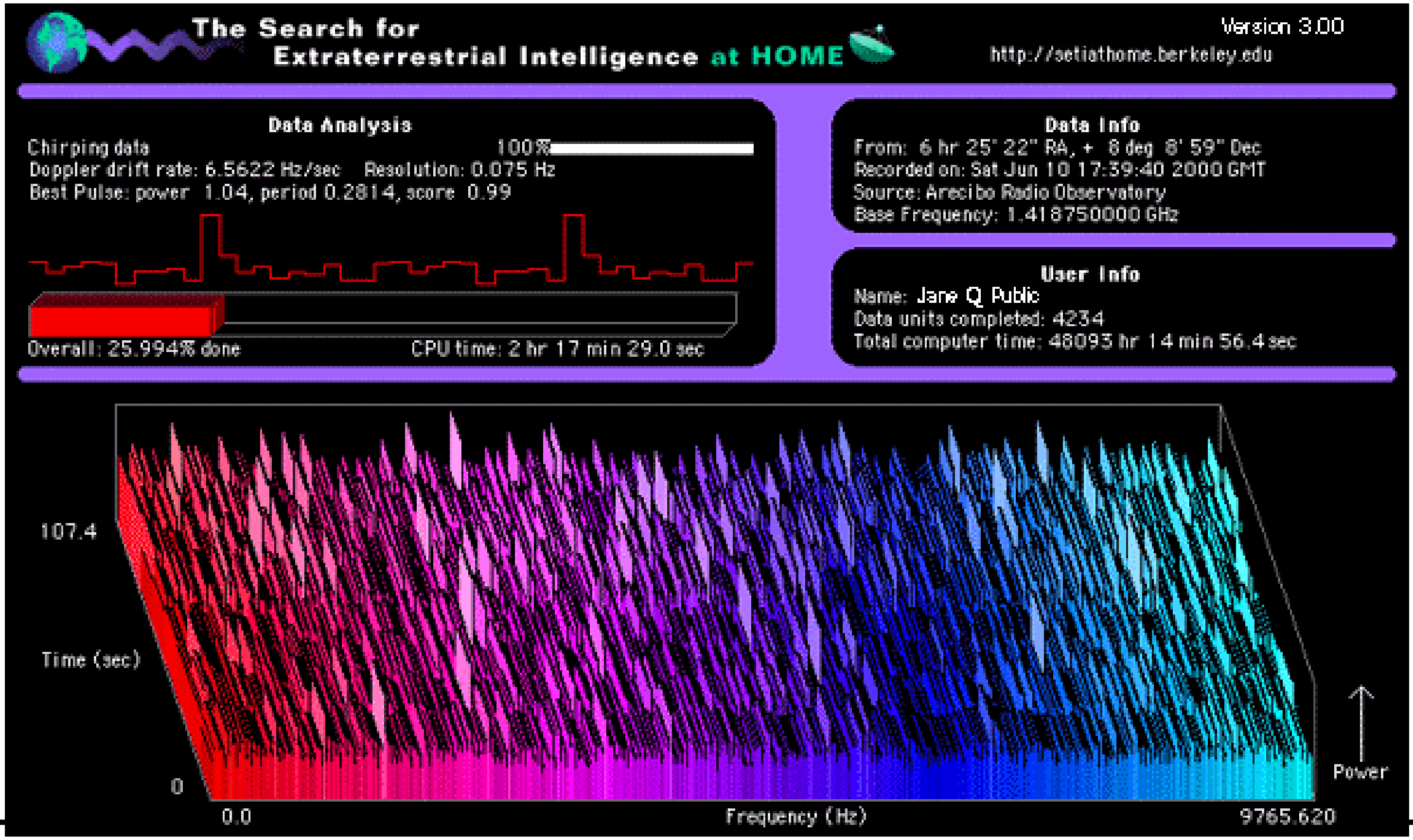


Fundamental Idea: Resource Sharing

- Resources for complex problems are distributed
  - Advanced scientific instruments (accelerators, telescopes, ...)
  - Storage and computing
  - Groups of people
- Communities require access to common services
  - Scientific collaborations (physics, astronomy, biology, eng. ...)
  - Government agencies
  - Health care organizations, large corporations, ...
- Goal is to build "Virtual Organizations"
  - Make all community resources available to any VO member
  - Leverage strengths at different institutions
  - Add people & resources dynamically

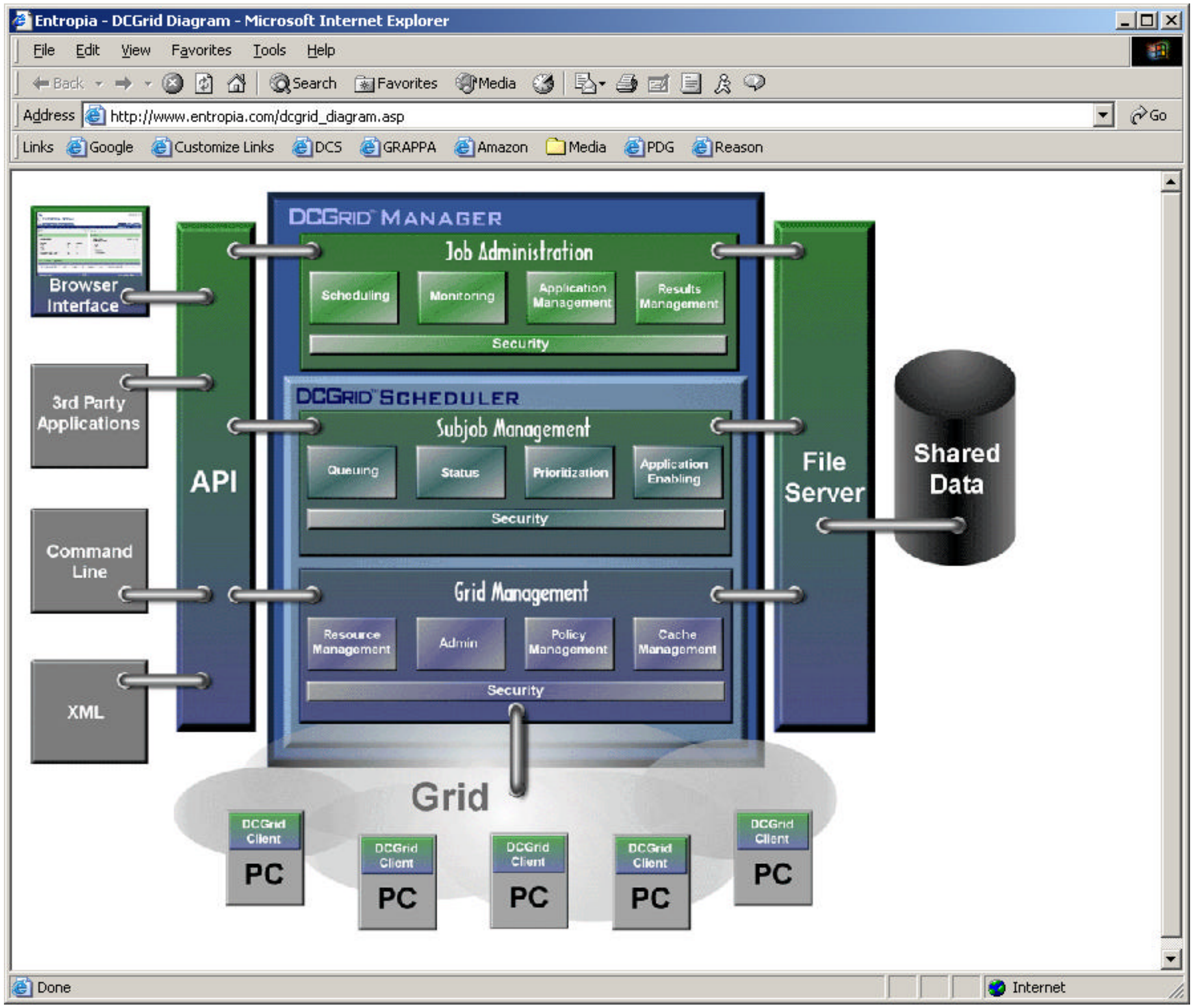
Proto-Grid: SETI@home

- ➔Community: SETI researchers + enthusiasts
- ➔Arecibo radio data sent to users (250KB data chunks)
- ➔Over 2M PCs used



More Advanced Proto-Grid: Evaluation of AIDS Drugs

- ➔Entropia
  - ◆DCGrid software
  - ◆Uses 1000s of PCs
- ➔Chief applications
  - ◆Drug design
  - ◆AIDS research



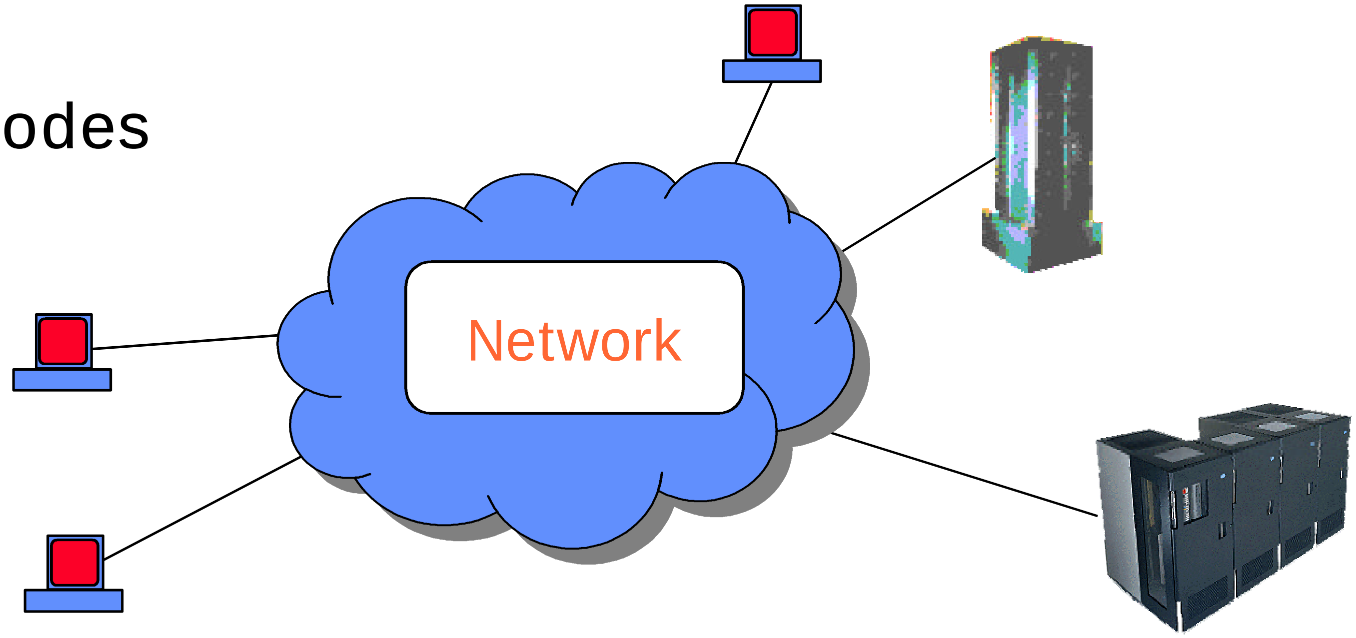
Some (Realistic) Grid Examples

- ➔ High energy physics
  - ◆ 3,000 physicists worldwide pool Petaflops of CPU resources to analyze Petabytes of data
- ➔ Climate modeling
  - ◆ Climate scientists visualize, annotate, & analyze Terabytes of simulation data
- ➔ Biology
  - ◆ A biochemist exploits 10,000 computers to screen 100,000 compounds in an hour
- ➔ Engineering
  - ◆ A multidisciplinary analysis in aerospace couples code and data in four companies to design a new airframe
- ➔ Many commercial applications

From Ian Foster

1990s Information Infrastructure

O(10^7) nodes

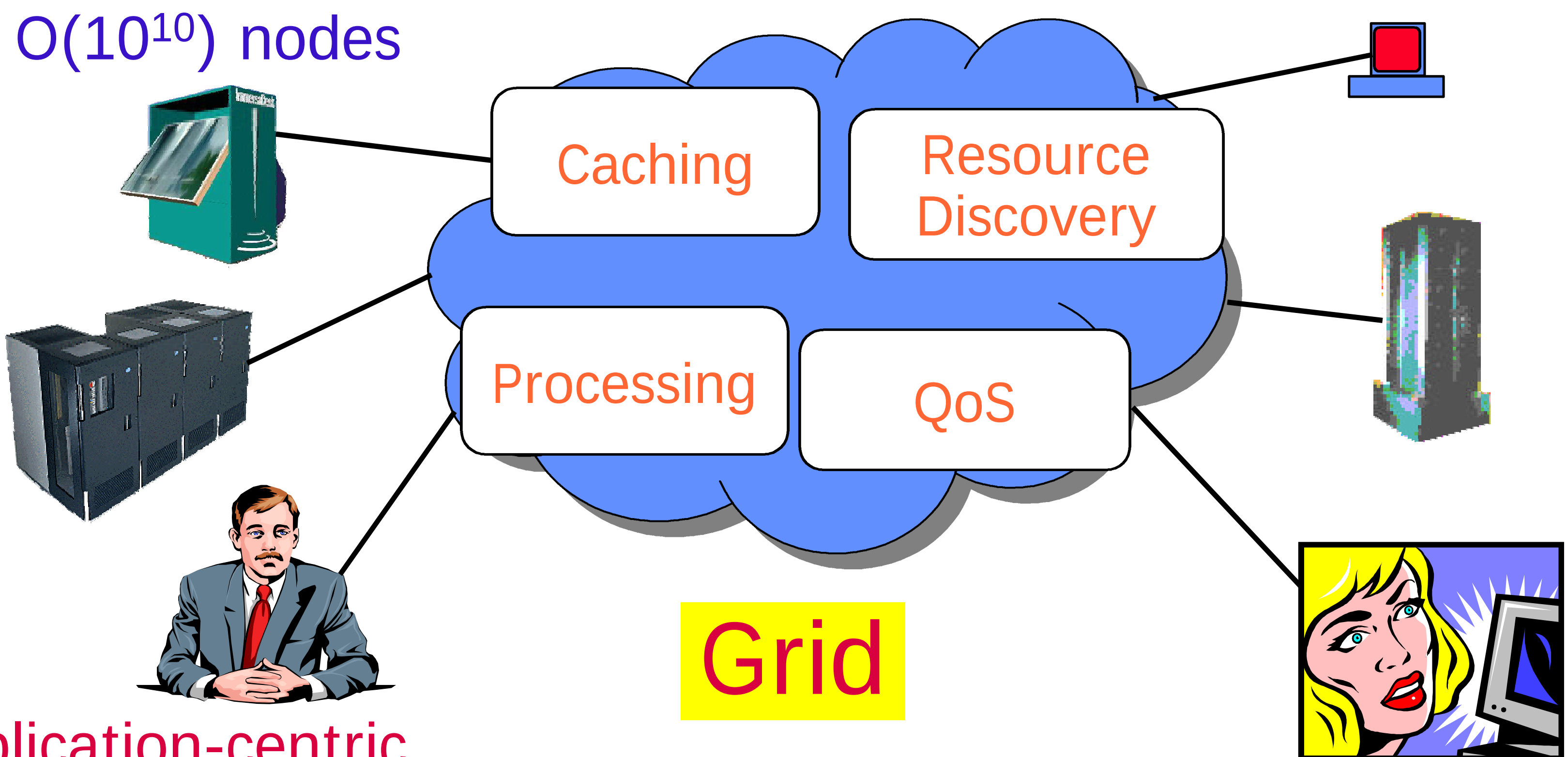


➔Network-centric

- Simple, fixed end systems
- Few embedded capabilities
- Few services
- No user-level quality of service

From Ian Foster

Emerging Information Infrastructure



➔Application-centric
- Heterogeneous, mobile end-systems
- Many embedded capabilities
- Rich services
- User-level quality of service

Qualitatively different, not just "faster and more reliable"

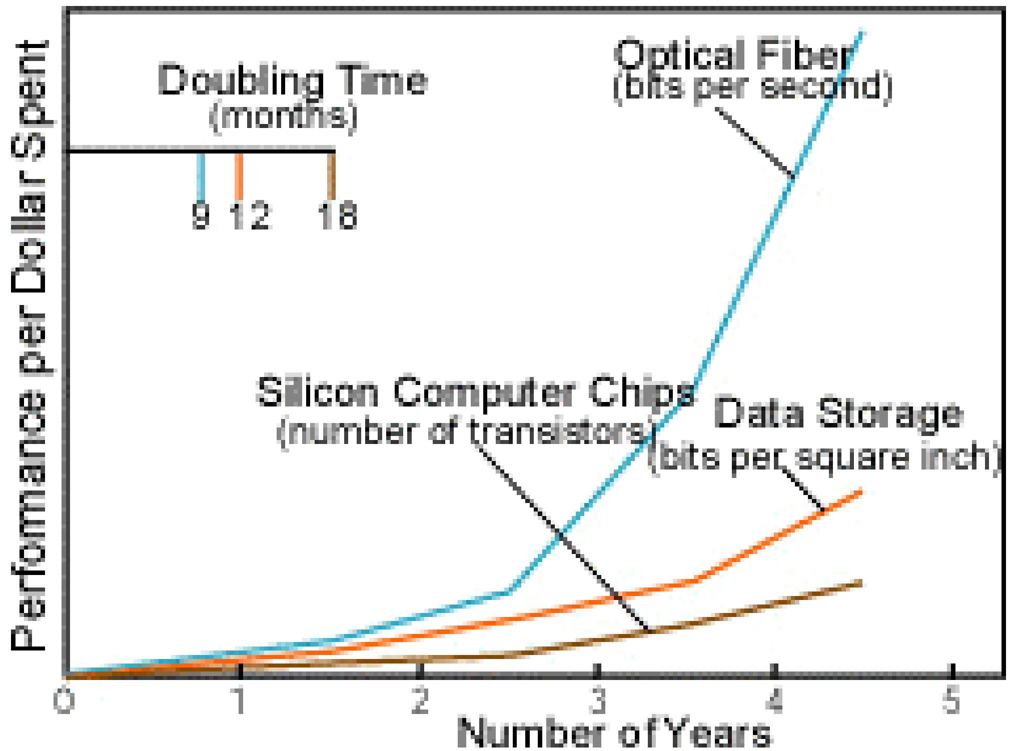
From Ian Foster

Grids: Why Now?

- ➔ Moore's law improvements in computing
  - ◆ Highly functional endsystems
- ➔ Universal wired and wireless Internet connections
  - ◆ Universal connectivity
- ➔ Changing modes of working and problem solving
  - ◆ Interdisciplinary teams
  - ◆ Computation and simulation as primary tools
- ➔ Network exponentials
  - ◆ (Next slide)

Network Exponentials & Collaboration

- ➔ Network vs. computer performance
  - ◆ Computer speed doubles every 18 months
  - ◆ Network speed doubles every 12 months (revised)
  - ◆ Difference = order of magnitude per 10 years
  - ◆ Other factor: network connectivity
- ➔ 1986 to 2001
  - ◆ Computers: × 1,000
  - ◆ Networks: × 50,000
- ➔ 2001 to 2010?
  - ◆ Computers: × 60
  - ◆ Networks: × 500



Scientific American (Jan-2001)

Evolution of Scientific Methodology

- ➔ Pre-electronic era
  - ◆ Theorize and/or experiment, alone or in small teams
  - ◆ Travel to conferences, publish papers, no cell phones ☺
- ➔ Transition period
  - ◆ Conduct team research at laboratories, data in diverse file formats
  - ◆ Laboratory holds all data, make data summaries for outside
  - ◆ Frequent travel to lab critical, ever ringing cell phones 😐
- ➔ Post-electronic era
  - ◆ Large international & global collaborations
  - ◆ Petabytes of observational, simulated data in distributed DBMS
  - ◆ Remote meetings, data exchanges & data viewing “instantaneous”
  - ◆ Remote operation of control rooms, test beams, etc.
  - ◆ Sit in nutrient bath, always connected, always communicating ☹

Basic Grid Challenges

- Overall goal: Coordinated sharing of resources
  - Resources under different administrative control
- Many technical problems to overcome
  - Authentication, authorization, policy, auditing
  - Resource discovery, access, negotiation, allocation, control
  - Dynamic formation & management of Virtual Organizations
  - Delivery of multiple levels of service
  - Autonomic management of resources
  - Failure detection & recovery
- Additional issue: lack of central control & knowledge
  - Preservation of local site autonomy

Advanced Grid Challenges

- ➔Goal: Secure workflow management & optimization
- ➔Maintaining a global view of resources and system state
  - ◆Coherent end-to-end system monitoring
  - ◆Adaptive learning: new paradigms for execution optimization
- ➔Workflow management
  - ◆Balance policy vs. instantaneous capability to complete tasks
  - ◆Balance effective resource use vs. fast turnaround for priority jobs
  - ◆Match resource usage to policy over the long term
  - ◆Goal-oriented algorithms: steering requests according to metrics
- ➔Handling user-Grid interactions: guidelines, agents
- ➔Building higher level services & integrated user env.

Layered Grid Architecture (Analogy to Internet Architecture)

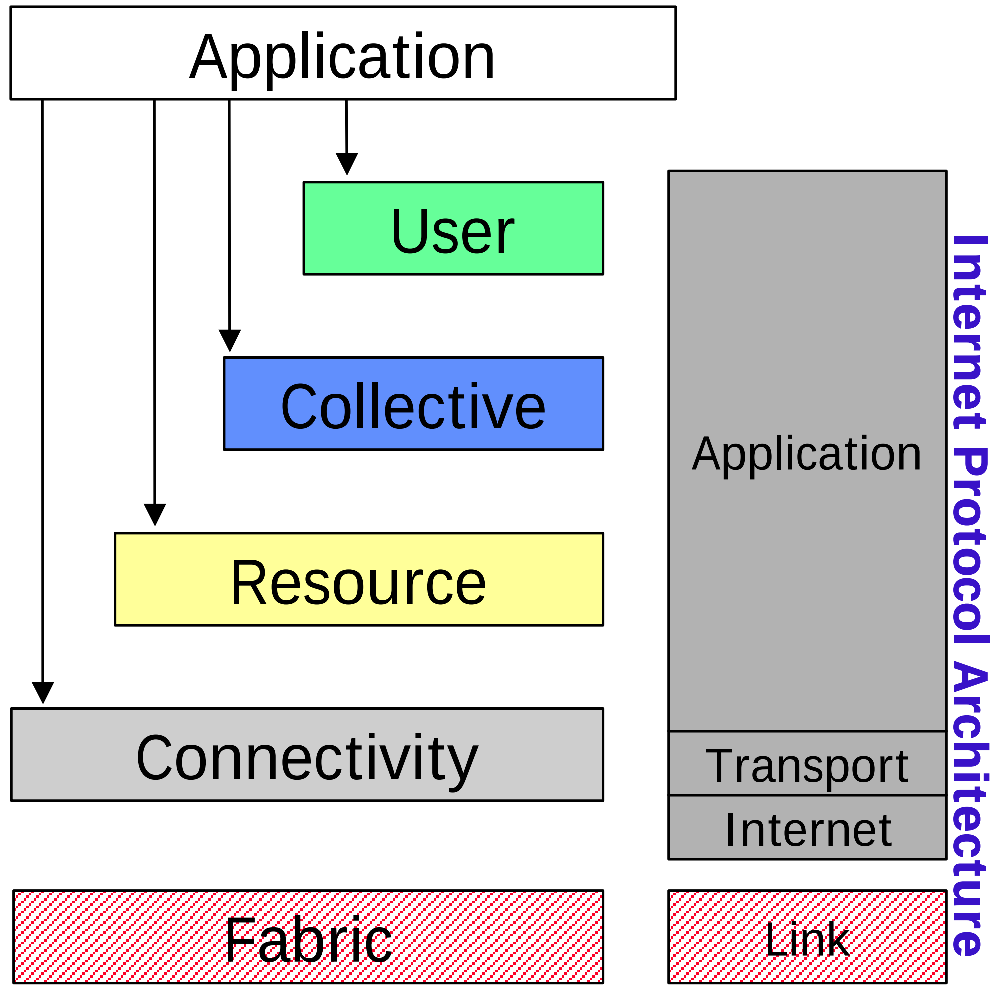
Specialized services:
App. specific distributed services

Managing multiple resources:
ubiquitous infrastructure services

Sharing single resources:
negotiating access, controlling use

Talking to things:
communications, security

Controlling things locally:
Accessing, controlling resources



From Ian Foster

Interfacing Software to the Grid

- ➔Physicist application codes
- ➔Expt. software framework layer: Modular, Grid aware
  - ◆Architecture able to interact effectively with the lower layers
- ➔Grid applications layer (for system operations)
  - ◆Policy and priority metrics
  - ◆Workflow evaluation metrics
  - ◆Task-site coupling proximity metrics
- ➔Global end-to-end system services layer
  - ◆Monitoring and tracking component performance
  - ◆Workflow monitoring and evaluation mechanisms
  - ◆Error recovery and redirection mechanisms
  - ◆System self-monitoring, evaluation & optimization mechanisms

Globus Project and Toolkit

- Globus Project™ (UC/Argonne + USC/ISI)
  - O(40) researchers & developers
  - Identify and define core protocols and services
- Globus Toolkit™ 2.0
  - Reference implementation of core protocols & services
  - Growing open source developer community
- Globus Toolkit used by most Data Grid projects today
  - US: GriPhyN, PPDG, TeraGrid, iVDGL, ...
  - EU: EU-DataGrid and national projects
- Recent announcement of applying “web services” to Grids
  - Keeps Grids in the commercial mainstream
  - GT 3.0

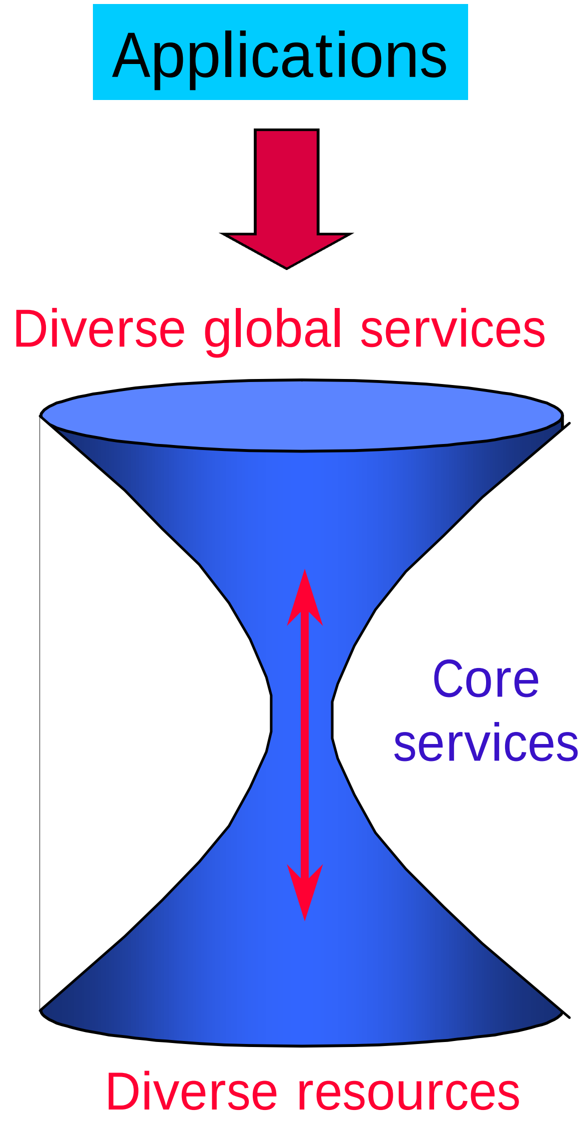
Globus General Approach

➔Define Grid protocols & APIs
- Protocol-mediated access to remote resources
- Integrate and extend existing standards

➔Develop reference implementation
- Open source Globus Toolkit
- Client & server SDKs, services, tools, etc.

➔Grid-enable wide variety of tools
- Globus Toolkit
- FTP, SSH, Condor, SRB, MPI, ...

➔Learn about real world problems
- Deployment
- Testing
- Applications



Data Grids

Data Intensive Science: 2000-2015

- Scientific discovery increasingly driven by IT
  - Computationally intensive analyses
  - Massive data collections
  - Data distributed across networks of varying capability
  - Geographically distributed collaboration
- Dominant factor: data growth (1 Petabyte = 1000 TB)
  - 2000 ~0.5 Petabyte
  - 2005 ~10 Petabytes
  - 2010 ~100 Petabytes
  - 2015 ~1000 Petabytes?

How to collect, manage, access and interpret this quantity of data?

Drives demand for "Data Grids" to handle additional dimension of data access & movement

Data Intensive Physical Sciences

- ➔ High energy & nuclear physics
  - ◆ Including new experiments at CERN's Large Hadron Collider
- ➔ Gravity wave searches
  - ◆ LIGO, GEO, VIRGO
- ➔ Astronomy: Digital sky surveys
  - ◆ Sloan Digital sky Survey, VISTA, other Gigapixel arrays
  - ◆ "Virtual" Observatories (multi-wavelength astronomy)
- ➔ Time-dependent 3-D systems (simulation & data)
  - ◆ Earth Observation, climate modeling
  - ◆ Geophysics, earthquake modeling
  - ◆ Fluids, aerodynamic design
  - ◆ Pollutant dispersal scenarios

Data Intensive Biology and Medicine

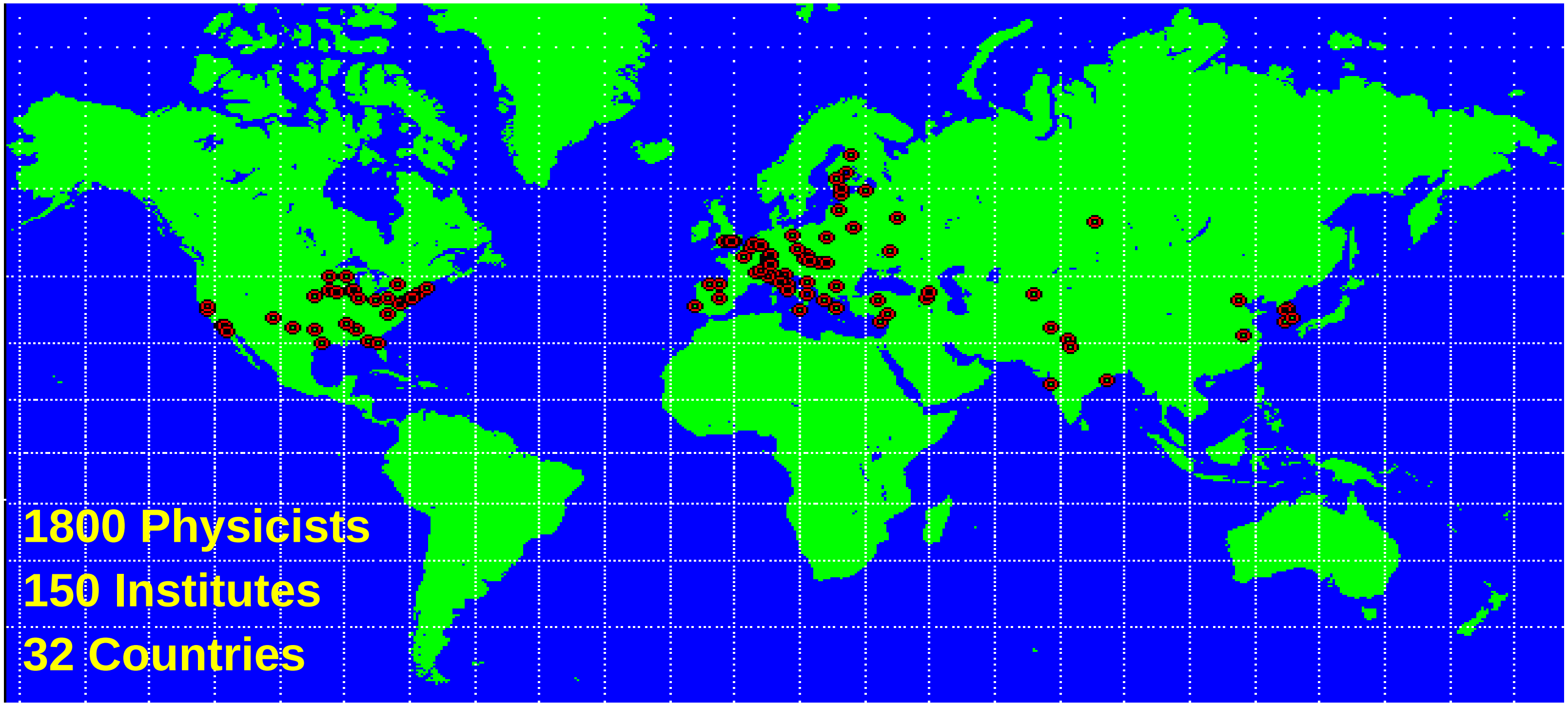
- Medical data
  - X-Ray, mammography data, etc. (many petabytes)
  - Digitizing patient records (ditto)
- X-ray crystallography
  - Bright X-Ray sources, e.g. Argonne Advanced Photon Source
- Molecular genomics and related disciplines
  - Human Genome, other genome databases
  - Proteomics (protein structure, activities, ...)
  - Protein interactions, drug delivery

Craig Venter keynote @SC2001

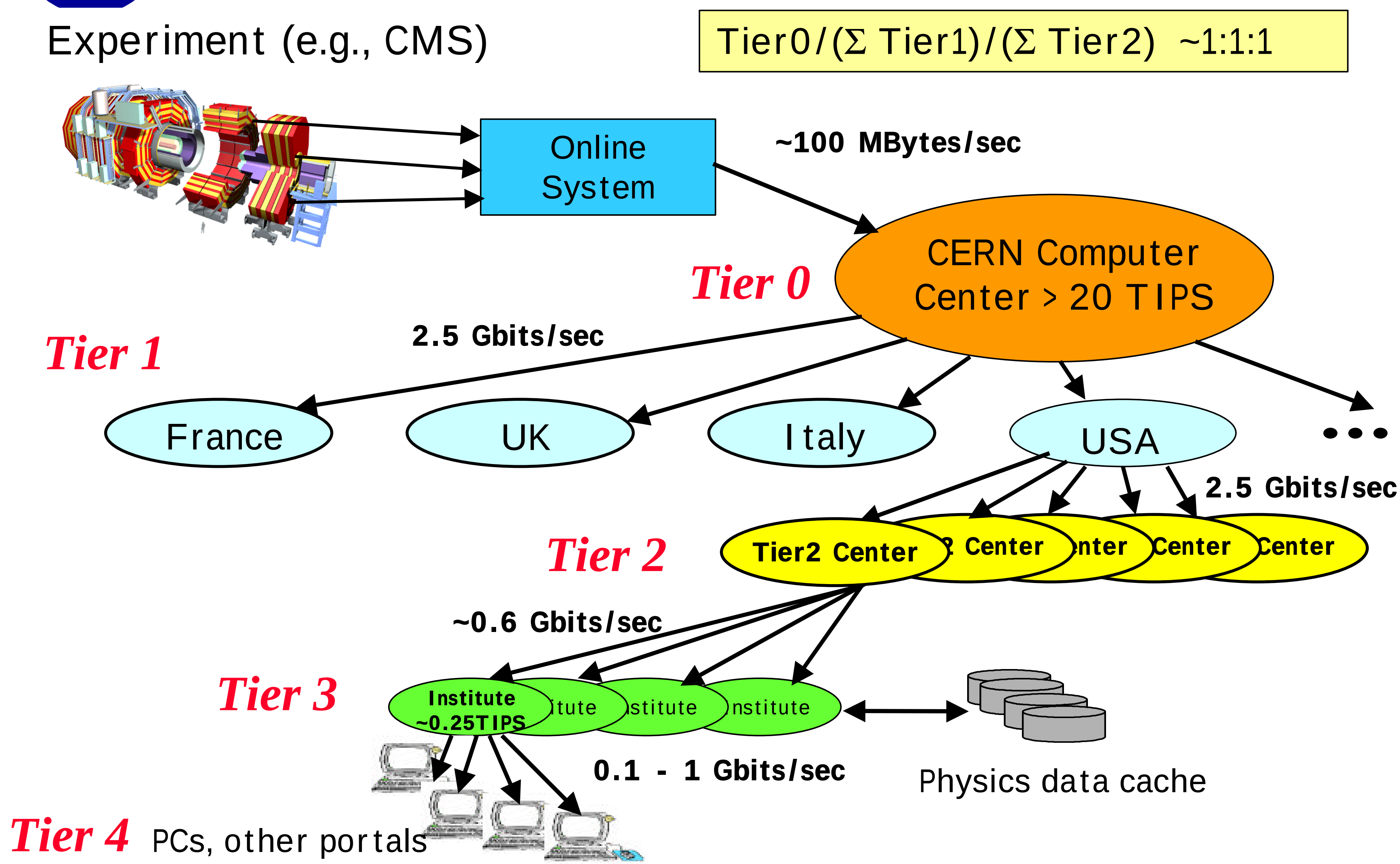
- Brain scans (3-D, time dependent)
- Virtual Population Laboratory (proposed)
  - Database of populations, geography, transportation corridors
  - Simulate likely spread of disease outbreaks

Example: LHC Computing Challenges

- ➔Complexity: Millions of individual detector channels
- ➔Scale: Petabytes of data / year (100 PB by ~2012)
- ➔Distribution: Global distribution of people & resources

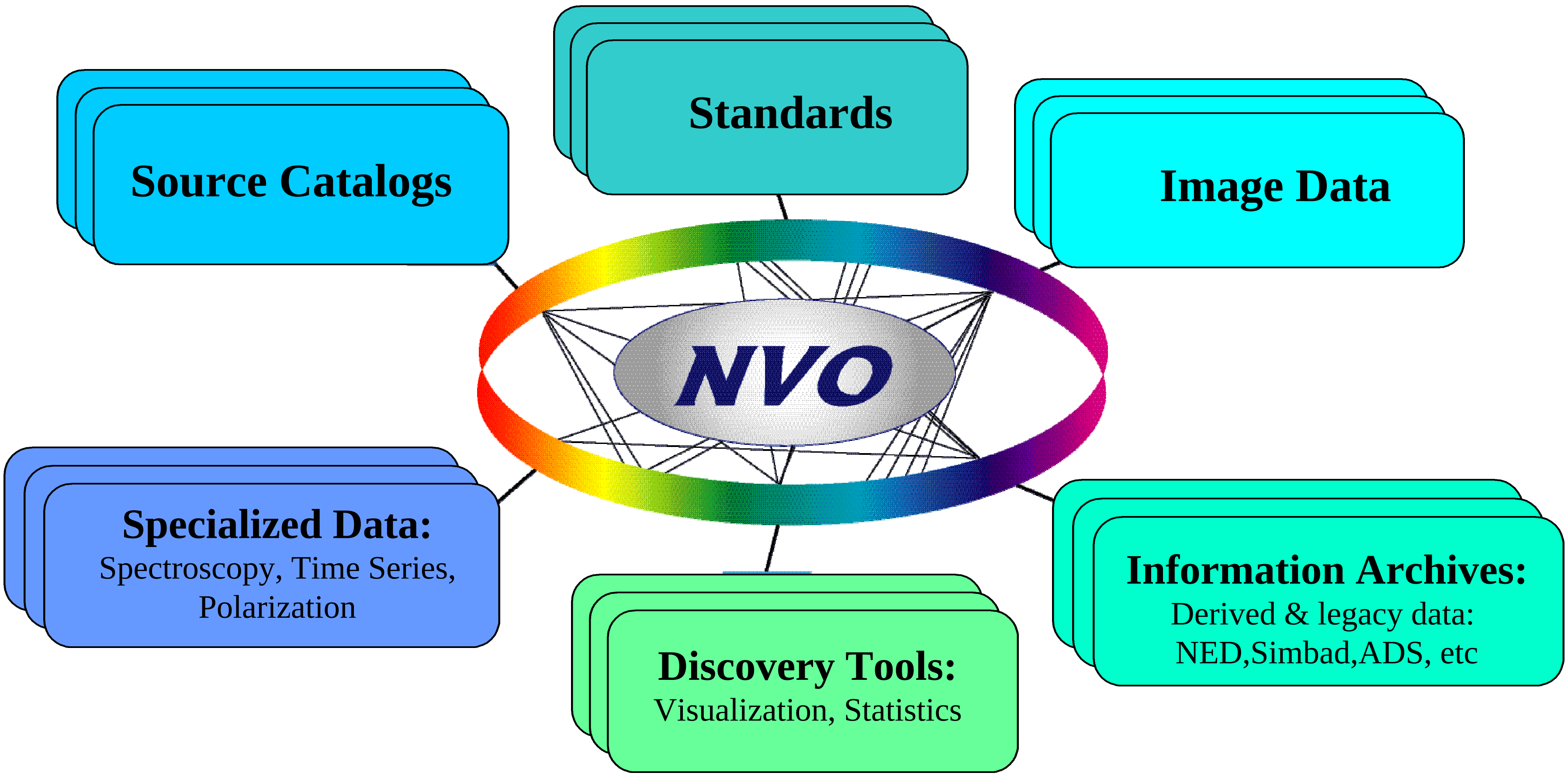


Example: Global LHC Data Grid

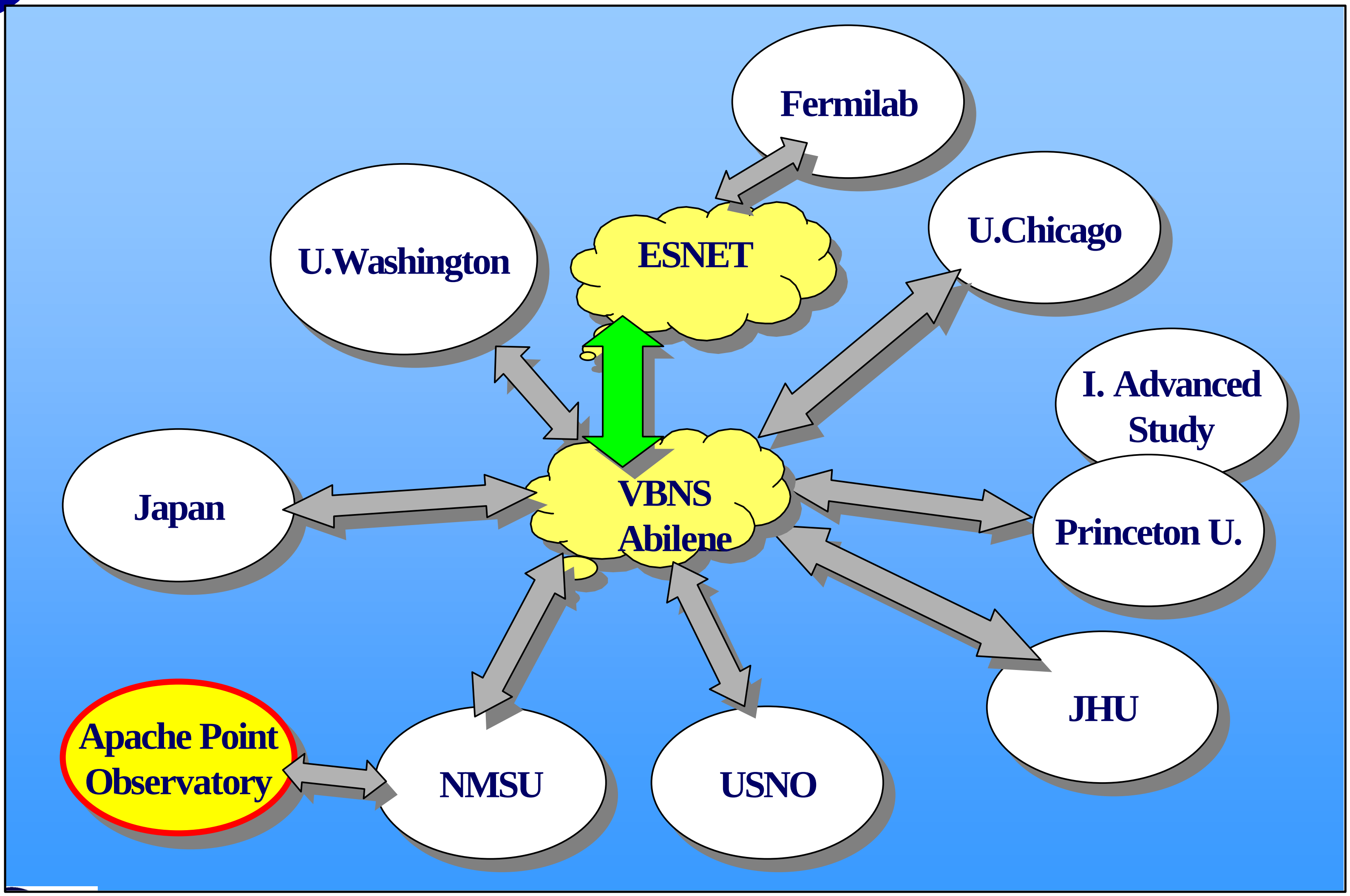


Example: Virtual Observatories

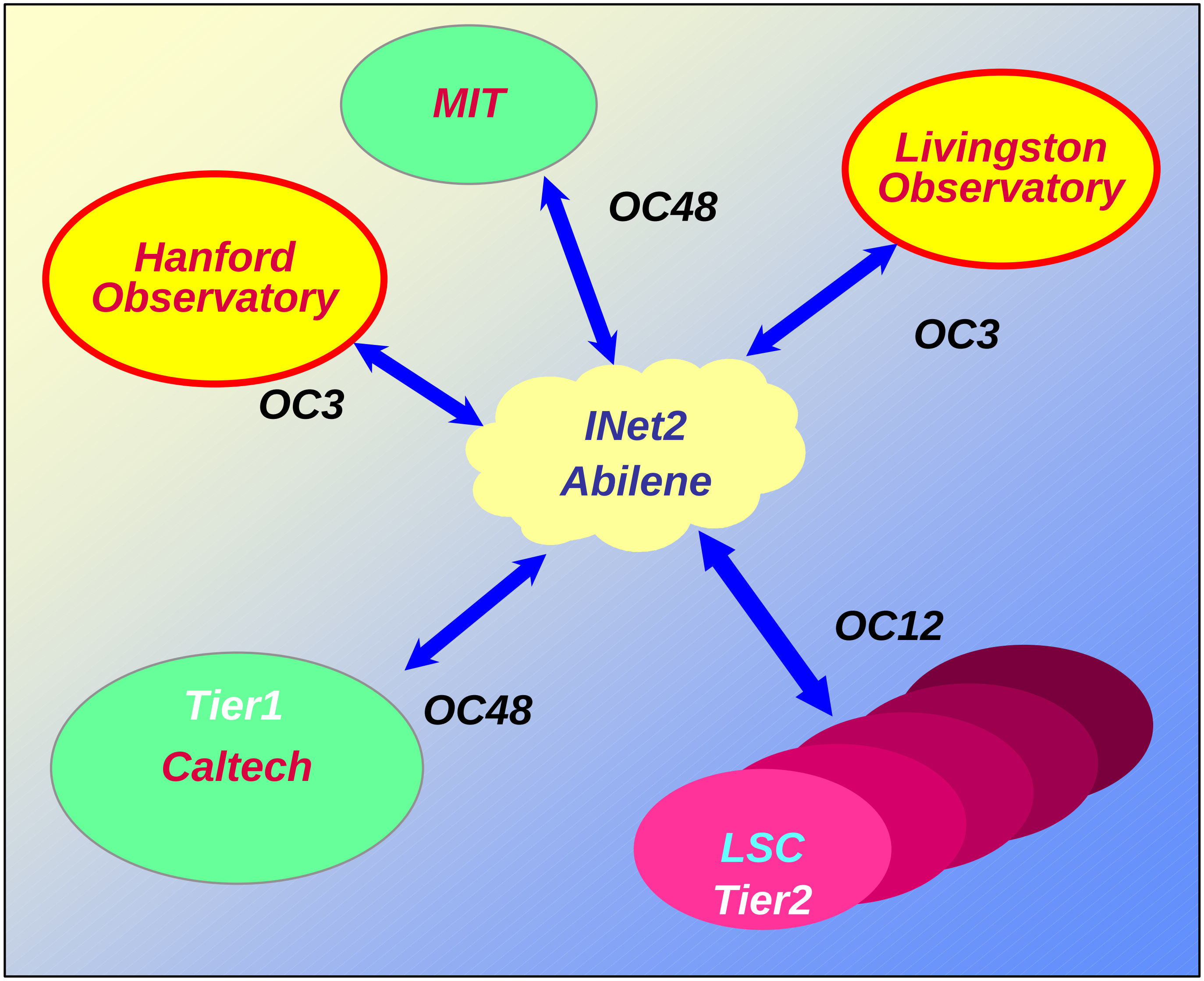
Multi-wavelength astronomy, Multiple surveys



Example: Sloan Sky Survey Data Grid



Example: LIGO Data Grid



Data Grid Projects

New Collaborative Endeavors via Grids

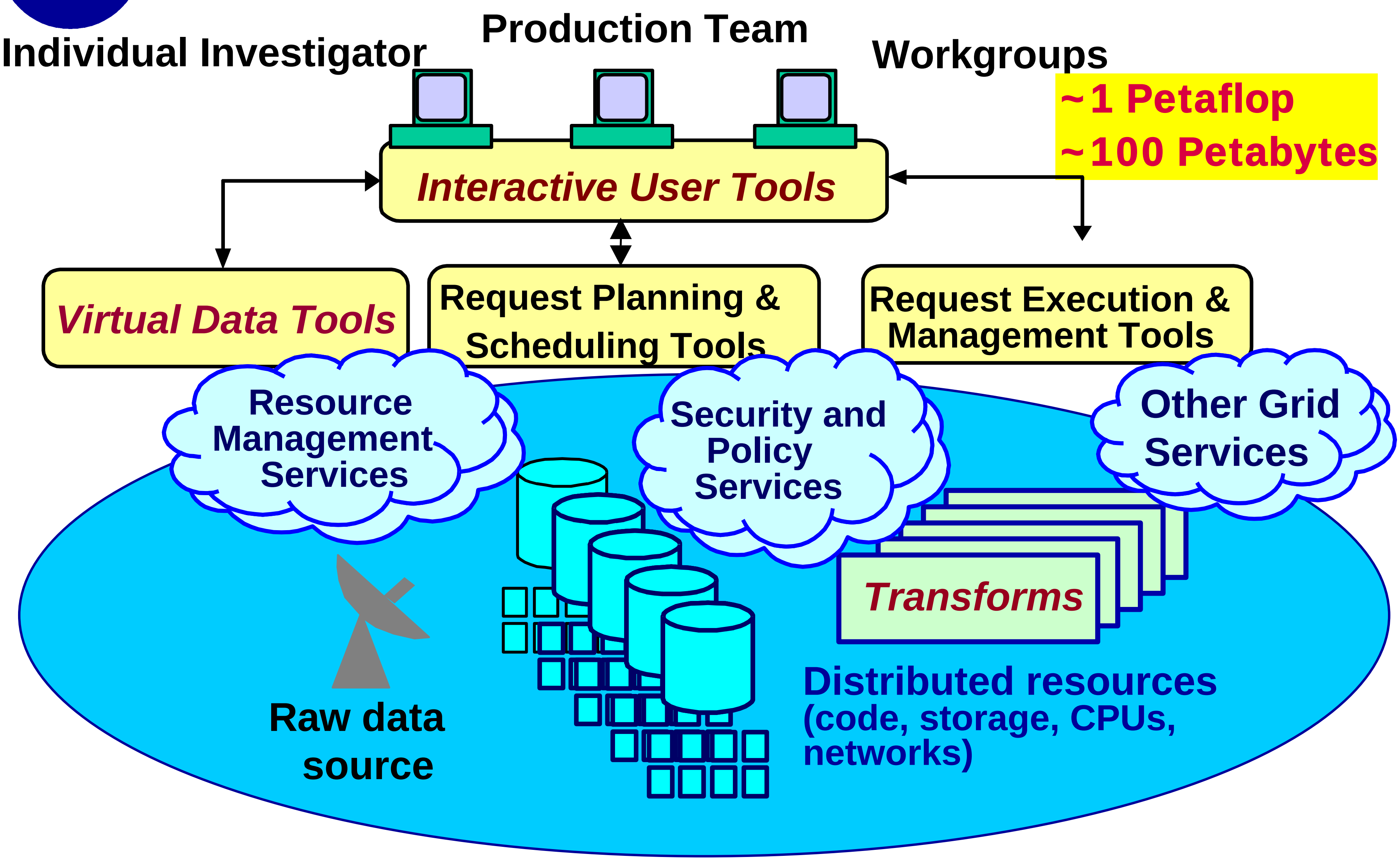
- Fundamentally alters conduct of scientific research
  - Old: People, resources flow inward to labs
  - New: Resources, data flow outward to universities
- Strengthens universities
  - Couples universities to data intensive science
  - Couples universities to national & international labs
  - Brings front-line research to students
  - Exploits intellectual resources of formerly isolated schools
  - Opens new opportunities for minority and women researchers
- Builds partnerships to drive new IT/science advances
  - Physics ⇔ Astronomy, biology, etc.
  - Application sciences ⇔ Computer Science
  - Universities ⇔ Laboratories
  - Fundamental sciences ⇔ IT infrastructure
  - Research Community ⇔ IT industry

Background: Major Data Grid Projects

- Particle Physics Data Grid (US, DOE)
  - Data Grid applications for HENP expts.
- GriPhyN (US, NSF)
  - Petascale Virtual-Data Grids
- iVDGL (US, NSF)
  - Global Grid lab
- DOE Science Grid (DOE)
  - Link major DOE computing sites
- TeraGrid (US, NSF)
  - Dist. supercomp. resources (13 TFlops)
- European Data Grid (EU, EC)
  - Data Grid technologies, EU deployment
- CrossGrid (EU, EC)
  - Realtime Grid tools
- DataTAG (EU, EC)
  - Transatlantic network, Grid applications
- Japanese Grid Project (APGrid?) (Japan)
  - Grid deployment throughout Japan

- Data intensive expts.
- Collaborations of application scientists & computer scientists
- Infrastructure devel. & deployment
- Globus based

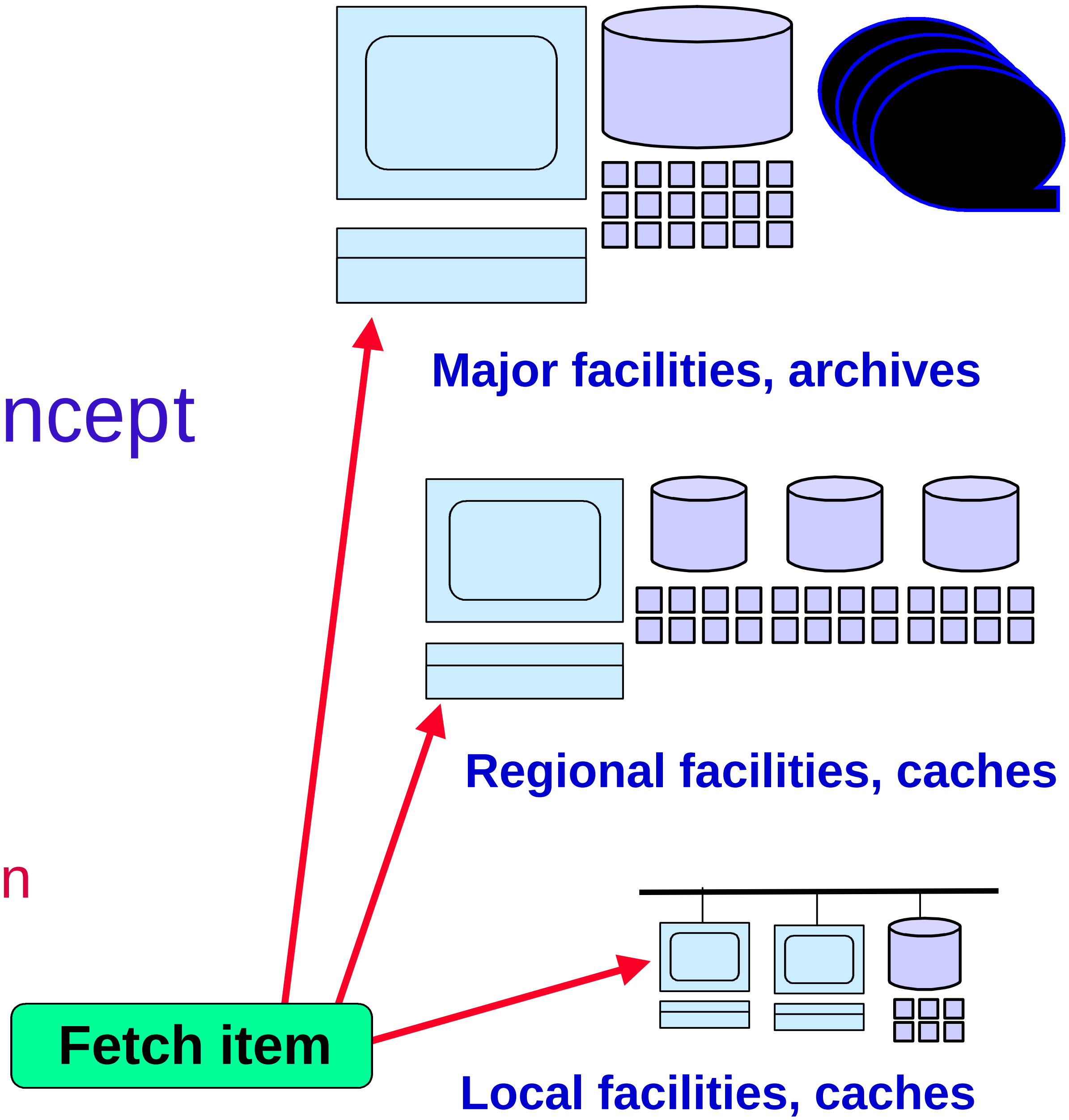
GriPhyN: PetaScale Virtual-Data Grids



GriPhyN
Data Intensive Science

Virtual Data Concept

- Data request may
  - Compute locally
  - Compute remotely
  - Access local data
  - Access remote data
- Scheduling based on
  - Local policies
  - Global policies
  - Cost



Particle Physics Data Grid

"PPDG will develop, evaluate and deliver vitally needed Grid-enabled tools for data-intensive collaboration in particle and nuclear physics. Novel mechanisms and policies will be vertically integrated with Grid Middleware, experiment specific applications and computing resources to provide effective end-to-end capability."

From PPDG proposal, 2001

Computer Science Program of Work

- **CS1: Job Description Language**
- **CS2: Schedule and Manage Data Processing and Placement Activities**
- **CS3 Monitoring and Status Reporting**
- **CS4 Storage Resource Management**
- **CS5 Reliable Replication Services**
- **CS6 File Transfer Services**
- **….**
- **CS11 Grid-enabled Analysis**

- Funded by DOE MICS ($9.5M for 2001-2004)
- DB replication, caching, catalogs
- Practical orientation: networks, instrumentation, monitoring

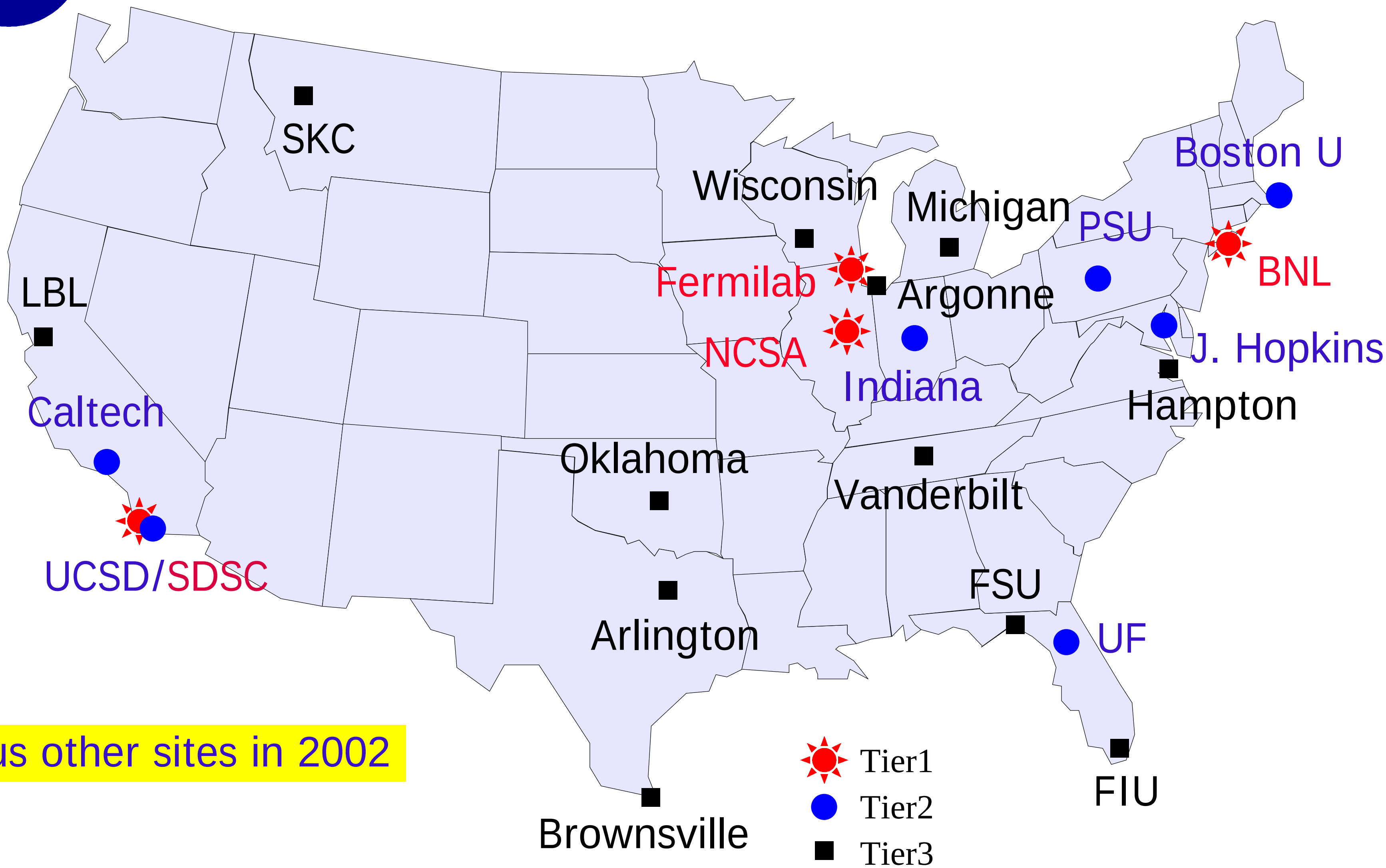
iVDGL: A Global Grid Laboratory

"We propose to create, operate and evaluate, over a sustained period of time, an international research laboratory for data-intensive science."

From NSF proposal, 2001

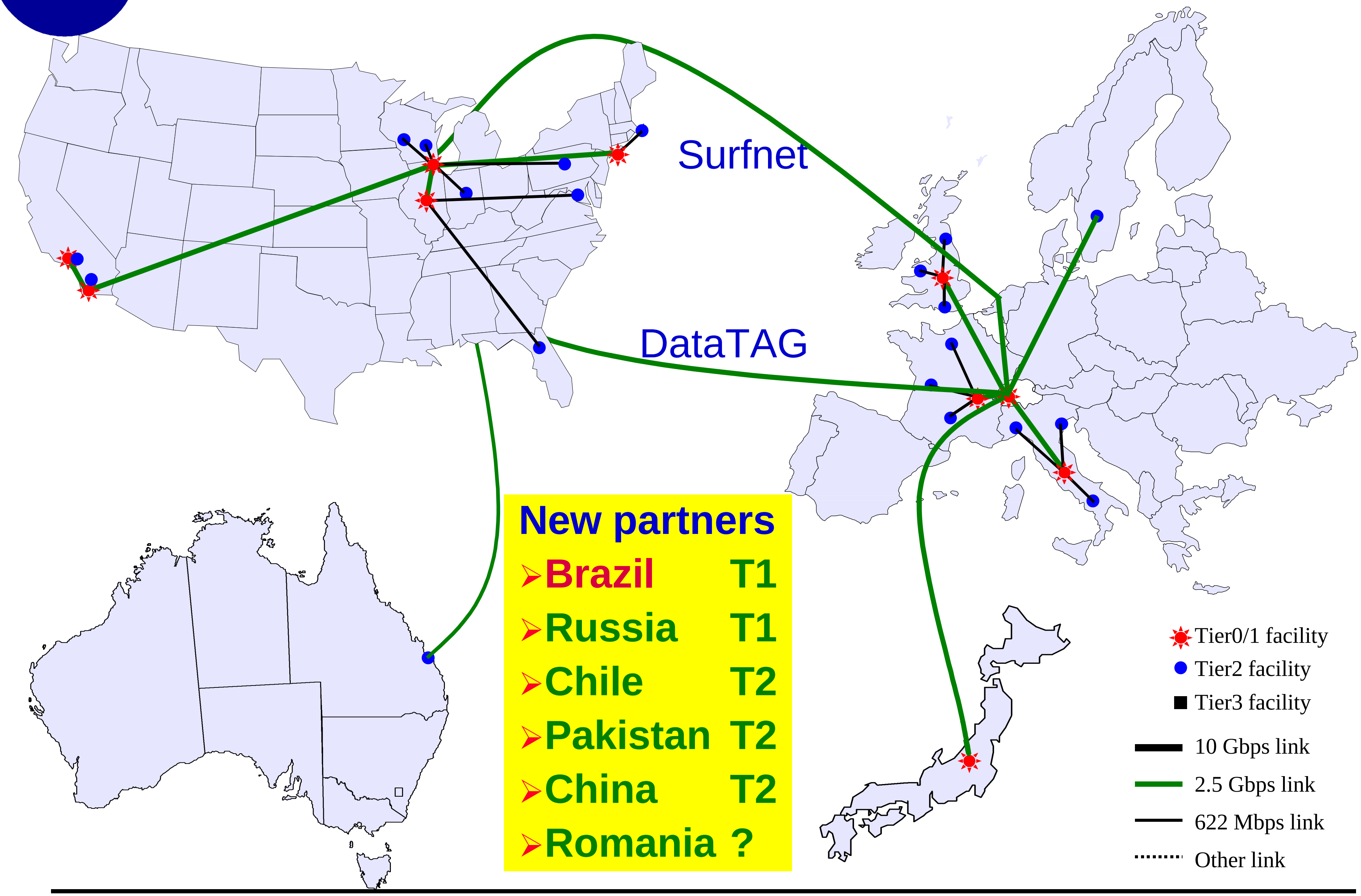
- International Virtual-Data Grid Laboratory
  - A global Grid laboratory (US, EU, Asia, South America, ...)
  - A place to conduct Data Grid tests "at scale"
  - A mechanism to create common Grid infrastructure
  - A laboratory for other disciplines to perform Data Grid tests
  - A focus of outreach efforts to small institutions
- U.S. part funded by NSF (2001-2006)
  - \$13.7M (NSF) + \$2M (matching)
  - UF directs this project
  - International partners bring own funds

US-iVDGL Data Grid (Dec. 2002)



Plus other sites in 2002

iVDGL Map (2002-2003)



New partners
- **Brazil T1**
- **Russia T1**
- **Chile T2**
- **Pakistan T2**
- **China T2**
- **Romania ?**

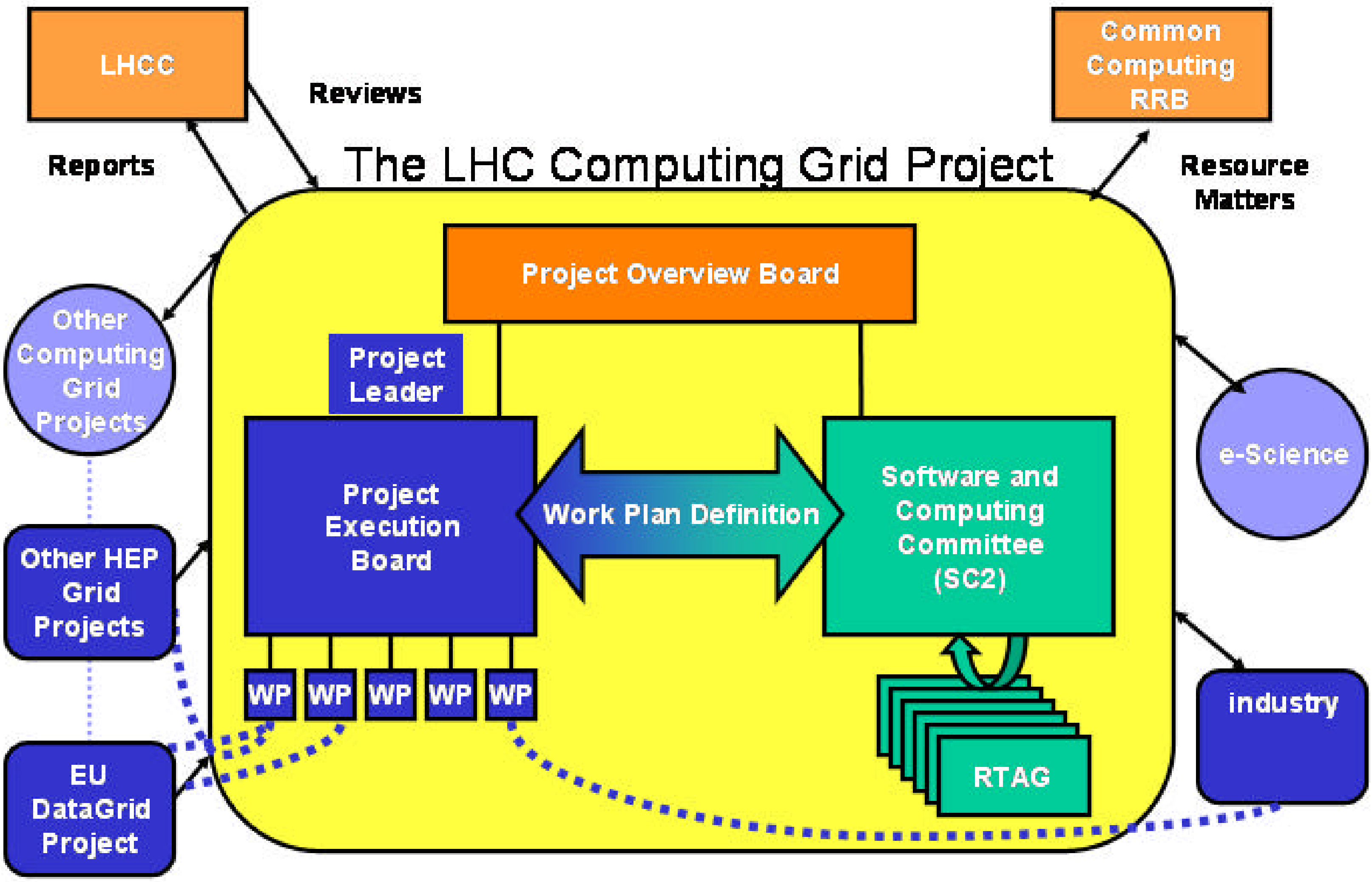
DOE Science Grid

➔Link major DOE computing sites (LBNL)

EU DataGrid Project

Work Package	Work Package title	Lead contractor
WP1	Grid Workload Management	INFN
WP2	Grid Data Management	CERN
WP3	Grid Monitoring Services	PPARC
WP4	Fabric Management	CERN
WP5	Mass Storage Management	PPARC
WP6	Integration Testbed	CNRS
WP7	Network Services	CNRS
WP8	High Energy Physics Applications	CERN
WP9	Earth Observation Science Applications	ESA
WP10	Biology Science Applications	INFN
WP11	Dissemination and Exploitation	INFN
WP12	Project Management	CERN

LHC Computing Grid Project



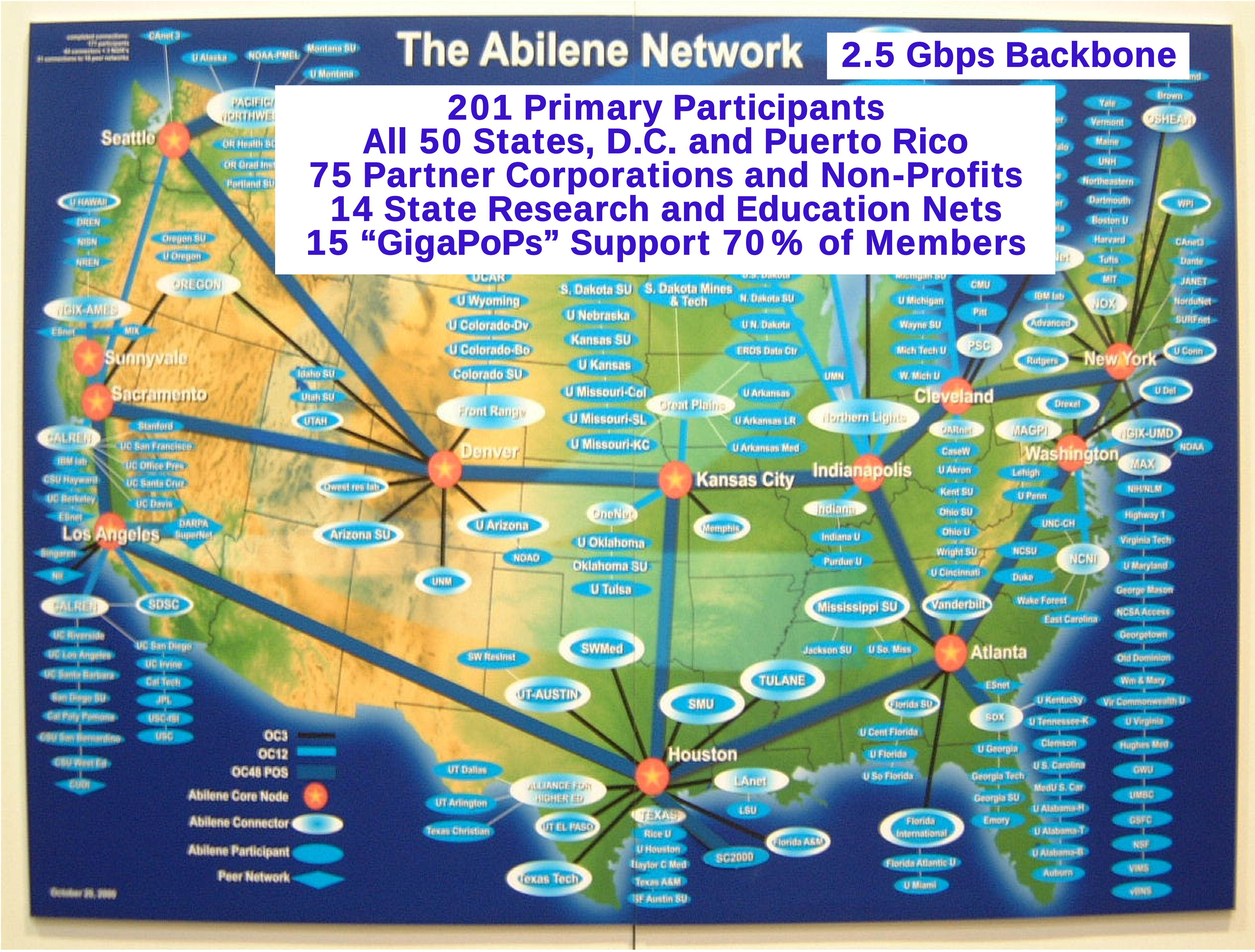
Networks

Next Generation Networks for HENP

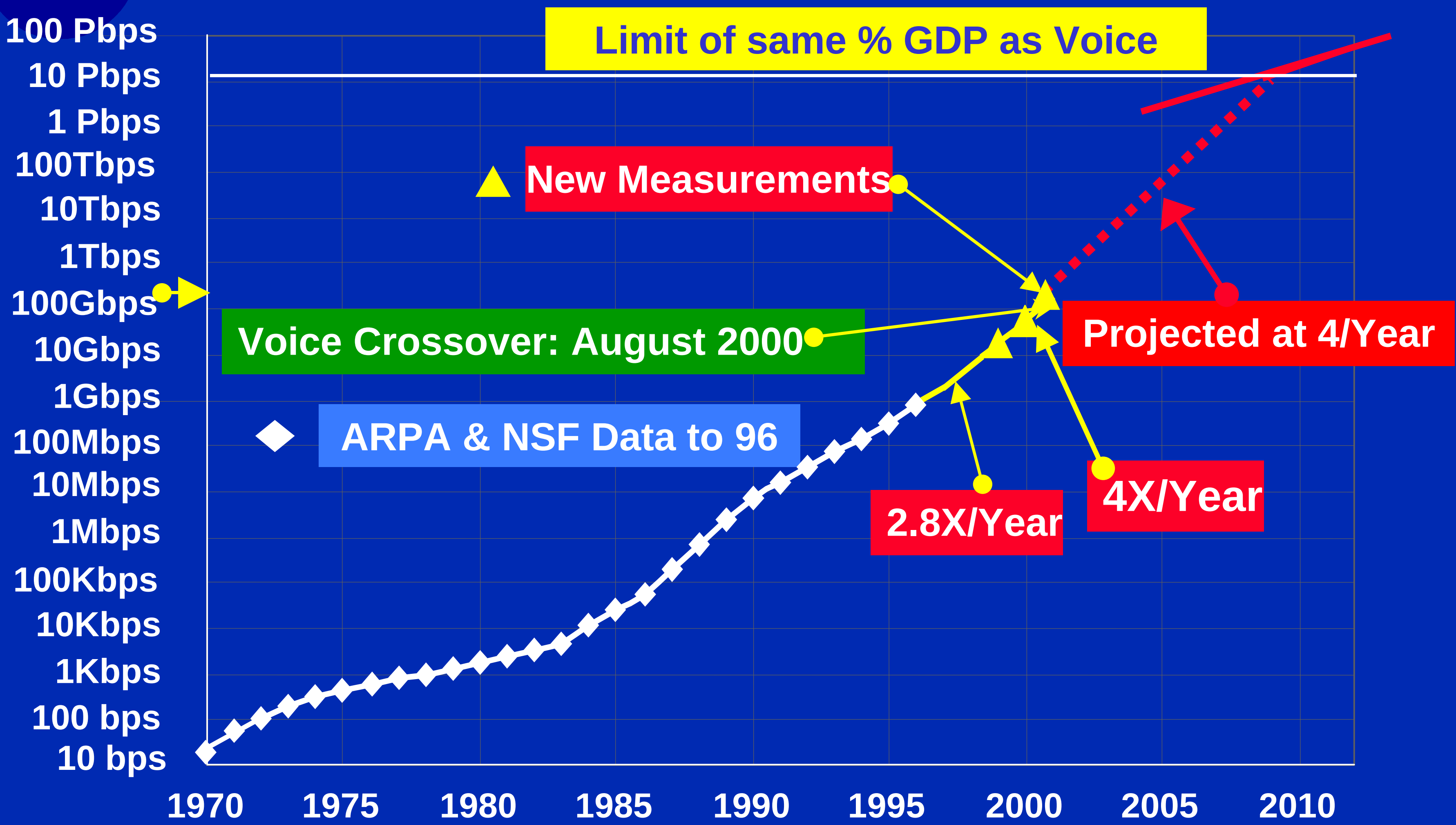
- ➔Rapid access to massive data stores
  - Petabytes and beyond
- ➔Balance of high throughput & rapid turnaround
  - Coordinate, manage computing, data, & network resources
- ➔Seamless high performance operation of WANs & LANs
  - Reliable, quantifiable (monitored), high performance
  - Rapid access to the data and computing resources
  - "Grid-enabled" data analysis, production and collaboration
- ➔Full participation by all physicists, regardless of location
  - Important requirement: good connectivity
  - Grid-enabled software, advanced networking, collaborative tools

2.5 Gbps Backbone

201 Primary Participants
All 50 States, D.C. and Puerto Rico
75 Partner Corporations and Non-Profits
14 State Research and Education Nets
15 “GigaPoPs” Support 70% of Members



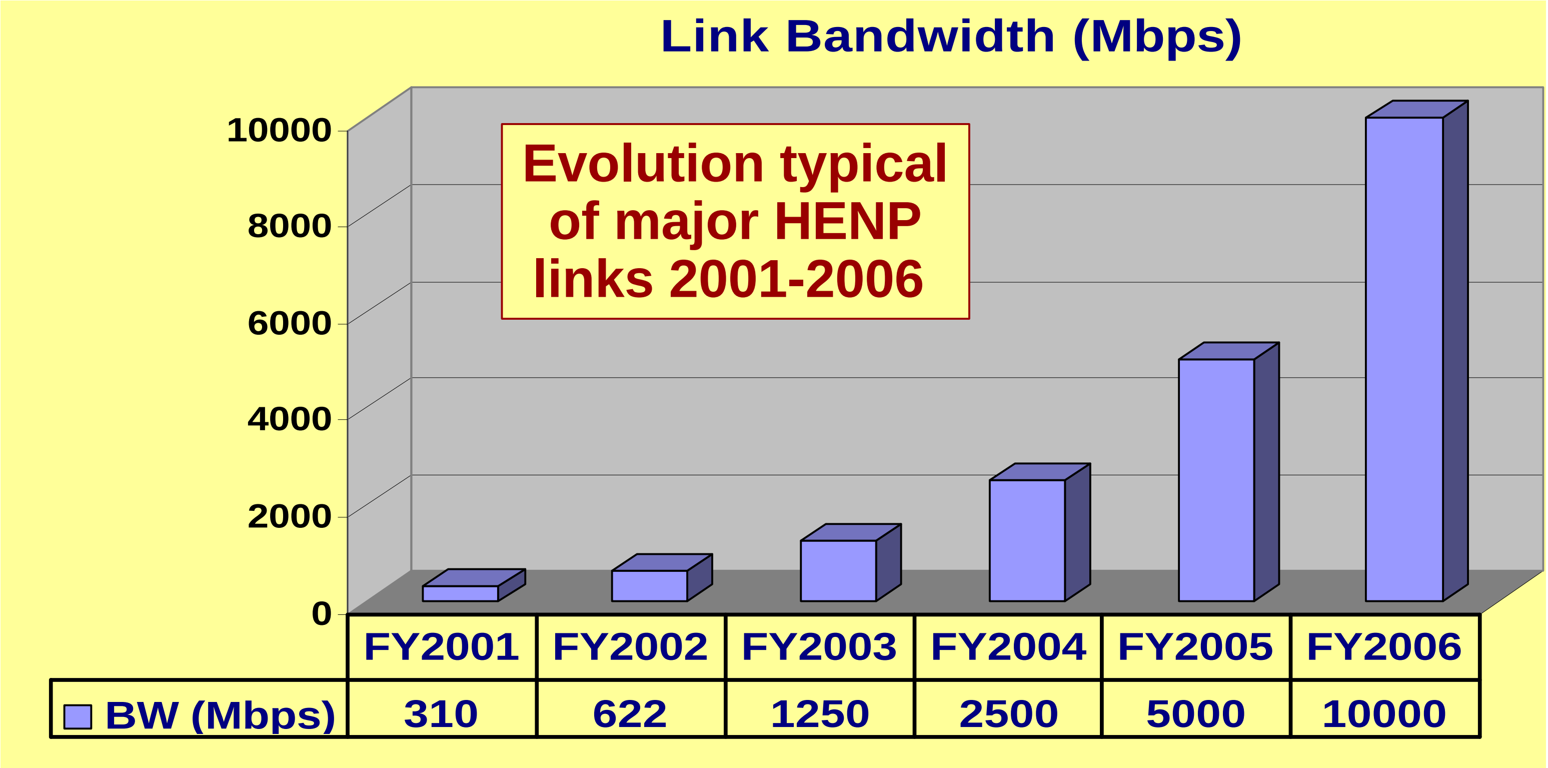
Total U.S. Internet Traffic



U.S. Internet Traffic

Source: Roberts et al., 2001

Baseline BW for the US-CERN Link: HENP Transatlantic WG: (DOE+NSF)



- 2×155 Mbps in 2001
- 622 Mbps May 2002
- 2.5 Gbps Research Link Summer 2002 (DataTAG)
- 10 Gbps Research Link in mid-2003 (DataTAG)

Transatlantic Network Working Group

	2001	2002	2003	2004	2005	2006
CMS	100	200	300	600	800	2500
ATLAS	50	100	300	600	800	2500
BaBar	300	600	1100	1600	2300	3000
CDF	100	300	400	2000	3000	6000
D0	400	1600	2400	3200	6400	8000
BTeV	20	40	100	200	300	500
DESY	100	180	210	240	270	300
CERN BW	155-310	622	2500	5000	10000	20000

- ➔BW in Mbps, assuming 50% utilization
- ➔See http://gate.hep.anl.gov/lprice/TAN

ICFA SCIC

- SCIC: Standing Committee on Interregional Connectivity
  - Created by ICFA in July 1998 in Vancouver
  - Make recommendations to ICFA concerning the connectivity between the Americas, Asia and Europe
- SCIC duties
  - Monitor traffic
  - Keep track of technology developments
  - Periodically review forecasts of future bandwidth needs
  - Provide early warning of potential problems
  - Create subcommittees when necessary
- Reports: February, July and October 2002

ICFA SCIC Topics

- Network status and upgrade plans
  - Bandwidth and performance evolution
  - Per country & transatlantic
- Performance measurements (world overview)
- Study specific topics
  - Example: Bulk transfer, VoIP, Collaborative Systems, QoS, Security
- Identification of problem areas
  - Ideas on how to improve, or encourage to improve
  - E.g., faster links $\Rightarrow$ equipment cost issues, TCP/IP scalability, etc.
- Last meeting in May 2002
  - Summary and sub-reports available
  - http://www.slac.stanford.edu/grp/scs/trip/notes-icfa-dec01-cottrell.html

Internet2 HENP Working Group

- Mission: Ensure the following HENP needs
  - National and international network infrastructures (end-to-end)
  - Standardized tools & facilities for high performance end-to-end monitoring and tracking
  - Collaborative systems
- Meet HENP needs in a timely manner
  - US LHC and other major HENP Programs
  - At-large scientific community
  - Create program broadly applicable across many fields
- Internet2 Working Group: Oct. 26 2001
  - Co-Chairs: S. McKee (Michigan), H. Newman (Caltech)
  - http://www.internet2.edu/henp (WG home page)
  - http://www.internet2.edu/e2e (end-to-end initiative)

All Major Links Advancing Rapidly

- ➔ Next generation 10 Gbps national network backbones
  - ◆ Starting to appear in the US, Europe and Japan
- ➔ Major transoceanic links
  - ◆ Are/will be at 2.5 - 10 Gbps in 2002-2003
- ➔ Critical path
  - ◆ Remove regional, last mile bottlenecks
  - ◆ Remove compromises in network quality
  - ◆ Prevent TCP/IP inefficiencies at high link speeds

U.S. Cyberinfrastructure Panel: Draft Recommendations

- New INITIATIVE to revolutionize science and engineering research at NSF and worldwide to capitalize on new computing and communications opportunities 21st Century Cyberinfrastructure includes supercomputing, but also massive storage, networking, software, collaboration, visualization, and human resources
  - Current centers (NCSA, SDSC, PSC) are a key resource for the INITIATIVE
  - Budget estimate: incremental $650 M/year (continuing)
- An INITIATIVE OFFICE with a highly placed, credible leader empowered to
  - Initiate competitive, discipline-driven path-breaking applications within NSF of cyberinfrastructure which contribute to the shared goals of the INITIATIVE
  - Coordinate policy and allocations across fields and projects. Participants across NSF directorates, Federal agencies, and international e-science
  - Develop high quality middleware and other software that is essential and special to scientific research
  - Manage individual computational, storage, and networking resources at least 100x larger than individual projects or universities can provide.

Grid References

- Grid Book
  - www.mkp.com/grids
- Globus
  - www.globus.org
- Global Grid Forum
  - www.gridforum.org
- TeraGrid
  - www.teragrid.org
- EU DataGrid
  - www.eu-datagrid.org
- PPDG
  - www.ppdg.net
- GriPhyN
  - www.griphyn.org
- iVDGL
  - www.ivdgl.org